Hindawi
Journal of Advanced Transportation
Volume 2024, Article ID 7311720, 15 pages
https://doi.org/10.1155/2024/7311720

Research Article

Metro Train Stopping Scheme Decision Based on Multisource Data in Express-Local Train Mode

Jin Li,[1] **Yaqiu Wang**,[1] **Shiyin Zhang**,[2] **and Huasheng Liu**[1]

[1]*College of Transportation, Jilin University, 5988 Renmin Street, Nanguan District, Changchun, China*
[2]*Public Security Bureau Traffic Police Detachment, Changchun, China*

Correspondence should be addressed to Huasheng Liu; liuhuasheng521@163.com

Received 20 September 2023; Revised 29 March 2024; Accepted 4 April 2024; Published 30 April 2024

Academic Editor: Luigi Dell'Olio

The urban rail transit network has gradually realized grid operation with the increase in the coverage rate. Therefore, the stopping schemes in accordance with the trend of the passenger flow are more conducive to improving the attractiveness of the rail transit and improving the sharing rate of the urban public transit. Traditional data from a single source may not be sufficient to describe the overall trend of the passenger flow in a period of time, and the error is possible in the case of insufficient data. Based on the multisource data, the spatial weight function is introduced to fuse the point of interest data and real estate information data, from which one obtains the residential index and office index, and the cluster analysis is conducted to obtain the potential stop scheme. Then, the optimization model of the train operation plan is established, aiming at minimizing the passenger travel time and the generalized system cost, and is constrained by a series of driving conditions. Compared with the single data source, multisource data can better reflect passenger flow trends and land use characteristics. Compared with the traditional all-station stopping scheme, a reasonable setting of crossing stations and running express-local trains can better satisfy the demands of the passenger flow. Finally, the optimization of Changchun rail Transit Line 1 shows that the model can reduce the travel time of passengers and the operating cost of the rail transit company and improve the quality of service, so as to achieve a win-win situation between passengers and the rail transit company.

1. Introduction

The conflict between the substantial travel demand of urban residents and the limited road traffic resources is intensifying and exacerbated by people's growing time consciousness. Congested road traffic conditions significantly undermine residents' travel experience. The subway has become the preferred mode of transportation for most residents due to its safety, efficiency, comfort, and speed. As a high-capacity means of transport, which does not occupy surface transportation resources, the subway has also become an integral component of transportation infrastructure development in major cities and is widely employed to alleviate pressure on ground transportation systems. Simultaneously, China's rapid urbanization necessitates continuous expansion of city boundaries and prompts the transition from a "single center" to a "multicenter" model. These developments and changes inevitably impose new requirements on public transportation systems and continue to rise long-haul passengers' travel demands steadily. The characteristics of passenger travel demand and flow distribution on long-distance rail transit lines differ significantly from those observed on existing rail transit lines. Compared with traditional urban rail transit station operations involving frequent stops, implementing fast and slow train stopping schemes can better cater to the travel needs of long-haul passengers.

To solve this problem, scholars at home and abroad have made the following typical studies on train operation schemes. Cury et al. [1] applied train operation schemes to the field of urban rail transit firstly. Before that, the research on train operation schemes was limited to the field of railway transportation, and the research content was the optimization of train numbers and departure intervals. Smrek [2] summarized the analytical techniques, operational models, and planning objectives applicable to the operation and organization of urban rail transit. Assis and Milani [3] proposed a method for calculating the optimal train schedule of subway lines based on a linear programming model, considering the time variation of passenger demand and line operating conditions. Wang et al. [4] considered various constraints in the actual train operation, established a multiobjective mixed integer nonlinear programming model, and obtained the dynamic train schedule and train turnover plan.

For the study of the operation scheme of the single interchange and cross-station parking mode, Niu et al. [5] proposed a quadratic integer programming model with linear constraints to meet the dynamic OD demand based on the cross-station stopping mode and minimize passenger waiting time and obtained the train schedule. Zhang et al. [6] analyzed time-varying passenger demand, proposed a flexible transit scheme that reduced the average travel time of passengers, and found the optimal scheme based on the genetic algorithm. Cao et al. [7] proposed a comprehensive evaluation model suitable for the express train mode, aiming at shortening the train running time and reducing the passenger waiting time and travel time, and the 0-1 integer programming model was established and solved by the Tabu algorithm and obtained the optimal train running scheme. Lee et al. [8] classified passengers according to starting and ending points and transfer choices, set up a coordination model of fast and local train modes with the goal of the shortest passenger travel time, designed a genetic algorithm, and obtained the optimal combination scheme of the fast and local train. Zhang et al. [9] proposed a timetable optimization model aiming at minimizing the total waiting time of passengers and solved it by genetic algorithm. Tang and Xu [10] established a double-layer programming model for the train operation scheme of suburban rail transit lines and designed a hybrid algorithm combining G-SA and MSA. Niu et al. [11] analyzed the problem of matching between the time-varying passenger flow demand and passenger waiting time and between trains and running routes in terms of the formulation and algorithm application of the train operation plan with the single route and all-station stopping mode, established a model with the goal of minimizing the passenger waiting time, and adopted the genetic algorithm. Deng et al. [12] constructed the elastic demand function of passenger travel, established a multiobjective double-layer programming model, and designed a simulated annealing algorithm. Based on IC card data, Zheng and Jin [13] proposed a new subway operation plan formulation method and obtained a departure plan that was closer to the passenger flow demand. Zhao et al. [14] established an integer linear programming model based on time-varying section requirements and predetermined service levels and used the two-stage method to solve the final train running plan. Based on AFC data, Yang et al. [15] proposed comprehensive optimization to formulate route plans and schedules. A mixed integer nonlinear programming model was established and solved by the improved NSGA-II.

In terms of the parallel operation of fast and local trains, Liu et al. [16] analyzed a line of rapid rail transit in Shanghai and discussed the reasonable crossing station location and siding setting. Sun et al. [17] established a double-layer programming model for fast and local train operation schemes. The upper layer model took the minimization of the passenger travel time and train turnover time as the target, and the lower layer model was the passenger flow allocation model, and it was solved by the particle swarm optimization algorithm. Zheng et al. [18] established a 0-1 integer programming model with the goal of maximizing the overall travel time of passengers under the cross-station parking mode and solved it by the Tabu search algorithm.

In summary, most studies primarily focus on designing subway operation schemes by designing operational routing, stopping schemes, optimization models, and model-solving algorithms. Most studies ignore the importance of input data, and the input data are often only single OD survey data or data obtained from the automated fare collection system (AFC) or prediction data derived from the aforementioned sources without considering the influence of land use and other factors on passenger flow trends. As a result, the designed optimization scheme is only suitable for the passenger flow in the input period of the model but not for the actual passenger flow generated by the line. Part of the reason is that data from single source cannot fully reflect the characteristics of the passenger flow. There is also a part of the factor that the passenger flow has a certain randomness and volatility, and the passenger flow data input to the model cannot capture the overall trend of passenger flows over a long period of time, and the smaller the amount of data, the greater the deviation. As urban centers continue to develop and grow in size, it has become increasingly evident that land use planning along subway stations significantly influences the passenger flow distribution, resulting in tidal passenger flows becoming a common occurrence. To address this issue caused by the uneven distribution of residential and employment lands, this study focuses on a specific subway line and proposes a train stopping scheme based on fast and local train modes. Multiple sources of data including AFC data, point of interest (POI) data, and real estate website data are collected for analysis purposes. Based on these datasets, an optimization model for the train stopping scheme is established with objectives of minimizing both the passenger travel time and system-generalized cost functions under various operational constraints. The NAGA-II algorithm with the elite strategy is employed. The proposed method is validated using Changchun Metro Line 1 as an illustrative example. The main contributions of this paper are as follows:

(1) Utilizing multisource data including the POI data, real estate website data, AFC data, and passenger survey questionnaire; a spatial weight function is introduced to obtain the land use index reflecting the distribution characteristics of the passenger flow. Based on these data, the express-local train stopping plan can be designed that adapts to the overall trend of the passenger flow without requiring extensive analysis or prediction.

(2) An optimization model for train stopping schemes is established by analyzing characteristics of the passenger flow distribution as well as passengers' choice behavior. This model determines the optimal proportion of fast and slow trains and maximizes limited resources while improving quality of service and travel efficiency.

The remainder of this paper is organized as follows. Section 2 provides detailed descriptions of issues studied, Section 3 establishes an optimization model for the train stopping scheme and designs the NSGA-II algorithm to solve it, Section 4 presents empirical research results regarding optimal ratios between fast and slow trains followed by discussion, and finally, Section 5 summarizes thesis findings while proposing future research directions.

2. Problem Description

2.1. Train Stopping Strategy. The train stopping plan specifies the stopping mode and stopping station. General urban rail transit train stops are mainly divided into two categories: one is all-station stopping mode and the other is the express-local train cross-station operation scheme. Local trains stop at every station to cater to the short-haul passenger flow and enhance operational accessibility. Express trains only stop at stations with high intensity of passenger flow, significantly reducing the travel time for long-haul passengers while also alleviating congestion on local trains as quickly as possible.

2.2. Classification of Passengers. From the point of passengers, they do not need to choose what kind of trains to take when running the all-station stopping mode. Passengers with different origins and destinations will choose the first train they encounter. When the express and local train modes are operated, passengers need to make a decision on whether to board the express train or the local train, and different decisions result in different travel times for passengers with the same origins and destinations. So, according to the origin and destination, passengers are divided into five categories: P1, P2, P3, P4, and P5. A is defined as the express station, and B is defined as the local station.

P1: The origin and destination are both express stations A, and both the express stations are adjacent.

P2: The origin and destination are both express stations A, and both the express stations are not adjacent.

P3: Passengers start at express station A and arrive at local station B.

P4: The pickup station is local station B, and the drop-off station is express station A.

P5: Passengers start and arrive at local station B.

The routes of the five categories of passengers are shown in Table 1.

For P3 and P4 passengers, if their origins and destinations are close (<5 km) or there are no transfer opportunities available, most of them will choose to take the local direct train. However, P3 long-haul passengers will board the express train when it arrives and then transfer if necessary. They may also opt to directly board a local train when one is available. Nevertheless, passengers often have low desire or demand for transfers. Thus, those with different types of departure and destination stations tend to choose the local direct train more frequently. So, P3, P4, and P5 can be classified into one category according to their choice behavior. As for P1, whether they take the express train or the local train, the on-board time is the same. However, for P2, different choice behaviors correspond to different on-board times, and the on-board time of choosing to take the express train will be significantly less than that of taking the local train. Therefore, P1 and P2 belong to two categories. With the previous one, there are three categories finally, L1, L2, and L3.

L1: Passengers whose departure and destination stations are both express stations, and both the express stations are adjacent.

L2: Passengers whose departure and destination stations are both express stations, and both the express stations are not adjacent.

L3: Local direct passengers.

2.3. Analysis of Passenger Travel Time. The travel time consists of the on-board time and the waiting time. The travel time is analyzed separately according to the classification of passengers, and the ratio of express to local trains is set as 1: n; the arrival of passengers is evenly distributed, so the average waiting time is half of the departure interval, the duration of a cycle is T, the pairs of the express train is D_f, and the pairs of the local train is D_s. For L1 and L2 passengers, their waiting time t_{wf} is expressed as follows:

$$t_{\mathrm{wf}} = \frac{1}{2}\frac{T}{D_f + D_s}. \quad (1)$$

For L3 passengers who choose a local train, the waiting time t_{ws} is

$$t_{\mathrm{ws}} = \frac{1}{2}\frac{T}{D_s}. \quad (2)$$

The on-board time includes three parts: the running time, the stopping time, and the start-stop additional time consumed during starting and braking. The on-board time t_{in} is given as follows:

$$t_{\mathrm{in}} = t_{\mathrm{od}} + \sum_i a_s \cdot \left(t_{\mathrm{stop}} + t_{qt}\right), \quad (3)$$

Table 1: Passenger's optional route.

Categories	Routes	
P1	A_1-A_2 (fast train)	A_1-A_2 (local train)
P2	A_1-A_2 (fast train)	A_1-A_2 (local train)
P3	A-transfer station-B (fast train)	A-B (local train)
P4	B-transfer station-A (fast train)	B-A (local train)
P5	B_1-transfer station 1-transfer station 2-B_2 (fast train)	B_1-B_2 (local train)

where t_{od} is the pure running time between the origin and destination; t_{stop} is the stopping time; t_{qt} is the start-stop additional time; a_s is 0-1 decision variable, if the train stops at s station, $a_s = 1$, otherwise, $a_s = 0$.

L1 passengers on-board time is given as follows:

$$t_{\text{in}}^{L1} = t_{\text{ij}} + \sum_{i=1}^{j-1}\left(t_{\text{stop}} + t_{\text{qt}}\right). \tag{4}$$

L2 passenger have two types of routes: express direct and local direct. Express direct t_{in}^{L21} is shown in formula (5) and local direct t_{in}^{L22} is shown in formula (6):

$$t_{\text{in}}^{L21} = t_{\text{ij}} + \sum_{i=1}^{j-1} a_s \cdot \left(t_{\text{stop}} + t_{\text{qt}}\right), \tag{5}$$

$$t_{\text{in}}^{L22} = t_{\text{ij}} + \sum_{i=1}^{j-1}\left(t_{\text{stop}} + t_{\text{qt}}\right). \tag{6}$$

L3 passengers on-board time t_{in}^{L3} is given as follows:

$$t_{\text{in}}^{L3} = t_{\text{ij}} + \sum_{i=1}^{j-1}\left(t_{\text{stop}} + t_{\text{qt}}\right). \tag{7}$$

3. Model Construction and Solution

3.1. Processing of Multisource Data. The stopping scheme for the express-local train is designed based on the multisource data. The data used in this paper include AFC data, POI data, real estate website data, and questionnaire data.

3.1.1. Preprocessing of AFC Data. The AFC system stores nine items of valid information on the passenger rail transit travel, including the date, ticket card number, station name, transaction type, credit card time, etc. By processing the relevant information, the spatial and temporal distribution characteristics of the passenger flow and the OD table of passengers joining the transfer can be obtained, shown in Table 2.

3.1.2. Amap POI Data. Each POI contains four aspects of information: name, category, coordinates, and land use nature. Considering the scope impact of the TOD mode, it takes the subway station as the center and the buffer range within European distance of 800 m of the surrounding as the research scope. The POI data are collected and sorted through Python to obtain the land use information near the station.

3.1.3. Real Estate Website Data. The data of "Fangtianxia" and "Lianjia201d" real estate websites are crawled through Python to obtain the distance, number of households, and floor area information of residential land and office land and to estimate the number of employees or residents near the site.

Match the real estate website data with the POI data, and input the information into the Excel if the information is successfully matched. Due to the limited information uploaded by the real estate website, there is a part of the property information missing. For POI with this situation, the floor area, number of buildings, and floor height of POI are recorded, and then, the number of households is estimated according to the height of each floor of 3 meters and each household of 100 square meters, and finally, the data are entered into the Excel.

For the estimation of the number of employees in office buildings, the building area can be used for calculation. The used area is calculated according to 60% of the building area, and the per capita office area is calculated according to 10 square meters to estimate the number of employees. which are given in Table 3.

3.1.4. Passenger Travel Characteristics Survey Questionnaire Data. Through a survey questionnaire, the basic travel characteristics data of subway passengers, such as the travel purpose, travel time, travel cost, and frequency of rail transit travel, are obtained.

The processing of multisource data involves data layer fusion based on the POI data and real estate information data, so as to obtain office and residence indices reflecting land use characteristics. Cluster analysis is then conducted on stations using the station passenger collector-distributor volume abstracted from the AFC data and land use index to determine the stopping mode, which belongs to the feature layer fusion. Finally, the train operation scheme model is established by inputting the stopping scheme and OD matrix, obtaining feasible schemes. Before determining the final train stopping plan, the survey questionnaire data and the potential stopping plan are integrated at the decision-making level. The data fusion process is shown in Figure 1.

3.2. Multisource Data Fusion Based on Spatial Weight Function. The spatial weight function ω in the geographical weighted regression model is introduced in the analysis of land use around the station, and the data from the real estate website are integrated. The principle is that the closer the things are, the closer they are to each other (Tobler's first law of geography), and the closer the distance, the greater the

Table 2: Passenger flow OD matrix in peak hours.

O	D														
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
A	0	26	19	348	343	271	751	822	216	112	139	350	192	35	1
B	0	11	7	227	262	298	725	976	200	117	108	272	156	30	0
C	0	0	6	127	130	216	589	841	209	75	100	213	173	23	0
D	0	0	0	31	9	165	277	776	162	79	124	242	231	56	0
E	0	0	0	0	6	23	68	371	76	65	47	131	121	27	4
F	0	0	0	0	0	4	35	198	49	37	43	119	105	16	0
G	0	0	0	0	0	0	11	258	70	133	147	393	268	67	0
H	0	0	0	0	0	0	0	465	481	418	415	1077	954	120	2
I	0	0	0	0	0	0	0	0	6	44	35	152	161	24	1
J	0	0	0	0	0	0	0	0	0	2	20	51	89	20	1
K	0	0	0	0	0	0	0	0	0	0	6	27	55	18	1
L	0	0	0	0	0	0	0	0	0	0	0	25	121	30	7
M	0	0	0	0	0	0	0	0	0	0	0	0	10	4	2
N	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0
O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table 3: POI data and real estate website data.

Station	Residence buildings	Distance (m)	Number of households	Office buildings	Distance (m)	Number of employees	Building area (m^2)
Huaqing road	RB 1	478	1104	OB 1	343	287	4783
	RB 2	751	994	OB 2	165	10691	178183
	RB 3	435	419	OB 3	289	4140	69000
				OB 4	226	3300	55000

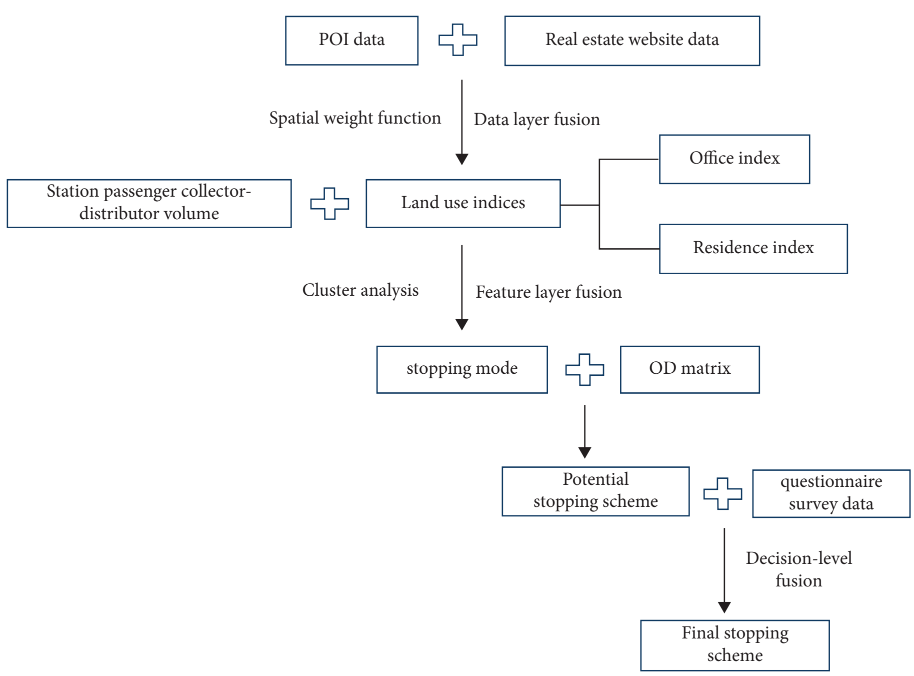

Figure 1: Multisource data fusion process.

spatial weight. After fusion, the residence index and office index of each station are output. The Gaussian kernel function is used to calculate the indices of various types of land use around the site. The residence index ω_{ilive} of site i is given as follows:

$$\omega_{\text{ilive}} = \sum_{j=1}^{n} p_{\text{ijlive}} \omega_{\text{ij}} = \sum p_{\text{ijlive}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{d_{\text{ij}}}{h} \right)^2 \right), \quad (8)$$

where p_{ijlive} is the number of households in the residential interest point ij, d_{ij} is the distance between i and j, and h is the window width

The office index ω_{iwork} of site i is given as follows:

$$\omega_{\text{iwork}} = \sum_{j=1}^{n} p_{\text{ijwork}} \omega_{\text{ij}} = \sum_{j=1}^{n} p_{\text{ijwork}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{d_{\text{ij}}}{h} \right)^2 \right), \quad (9)$$

where p_{ijwork} is the number of employees in the office interest point ij.

The standardized residence index, office index, and passenger flow data are given in Table 4. Stations 121, 128, and 135 are the terminal stations and transfer stations, which must be the stopping stations. To eliminate the interference of abnormal data, the data of stations 121, 128, and 135 are cleared.

3.3. Model Establishment

3.3. Model Establishment. The research focuses on a bi-directional single rail transit line without siding tracks during peak hours, and turn-back of trains is not considered. A multiobjective optimization model is established to minimize the total travel time cost of passengers and the operating cost of enterprises. Considering a series of driving conditions, the optimal pairs of express trains and local trains, and the stopping sequence of express trains are obtained. The parameters and variables involved in the model are given in Tables 5 and 6.

The assumptions of the model are as follows:

(1) All the passengers between express stations take express trains when available; otherwise, they will take local trains.
(2) Passengers between local stations choose local trains directly.
(3) The arrival of passengers is evenly distributed, and there are no stranded passengers at the station.
(4) The pairs of trains remain unchanged, and the formation of express and local trains is the same.

3.3.1. Objective 1. Minimum total passenger travel time cost

Total passenger travel time cost is the product of unit time value and passenger travel time

$$\min C_T = C_0 \cdot T. \quad (10)$$

The passenger travel time includes the passenger waiting time and on-board time

$$T = T_w + T_{\text{in}}. \quad (11)$$

According to the passenger classification method, the number of different types of passengers is determined as follows.

The number of L1 passengers

$$q_1 = \sum_{i=1}^{N-1} \sum_{j=2}^{N} a_i a_j q_{\text{ij}}. \quad (12)$$

The number of L2 passengers

$$q_2 = \sum_{i=1}^{N-1} \sum_{j=2}^{N} a_i \left(1 - a_j\right) q_{\text{ij}}. \quad (13)$$

The number of L3 passengers

$$q_3 = \sum_{i=1}^{N-1} \sum_{j=2}^{N} q_{\text{ij}}. \quad (14)$$

3.3.2. Waiting Time. L1 passengers waiting time

$$T_{w1} = \sum_{i=1}^{N-1} \sum_{j=2}^{N} a_i a_j q_{\text{ij}} t_{\text{wf}}. \quad (15)$$

L2 passengers waiting time

$$T_{w2} = \sum_{i=1}^{N-1} \sum_{j=2}^{N} a_i \left(1 - a_j\right) q_{\text{ij}} t_{\text{wf}} + \sum_{i=1}^{N-1} \sum_{j=2}^{N} \left(1 - a_i\right)\left(1 - a_j\right) q_{\text{ij}} t_{\text{wf}}. \quad (16)$$

L3 passengers waiting time

$$T_{w3} = \sum_{i=1}^{N-1} \sum_{j=2}^{N} q_{\text{ij}} t_{\text{ws}}. \quad (17)$$

Total passenger waiting time

$$T_w = q_1 T_{w1} + q_2 T_{w2} + q_3 T_{w3}. \quad (18)$$

3.3.3. On-Board Time. L1 passengers

$$T_{\text{in1}} = \sum_{i=1}^{N-1} \sum_{j=2}^{N} q_{\text{ij}} t_{\text{in}}^{L1}. \quad (19)$$

L2 passengers

$$f_{21} = \frac{D_f}{D_f + D_s}, \quad (20)$$

$$f_{22} = \frac{D_s}{D_f + D_s}, \quad (21)$$

$$T_{\text{in2}} = f_{21} \sum_{i=1}^{N-1} \sum_{j=2}^{N} q_{\text{ij}} t_{\text{in}}^{L21} + f_{22} \sum_{i=1}^{N-1} \sum_{j=2}^{N} q_{\text{ij}} t_{\text{in}}^{L22}. \quad (22)$$

L3 passengers

Table 4: Standardized data.

Station number	Upward	Downward	Residence index	Office index
122	1.23163	−0.94904	1.86879	−1.23794
123	0.55044	−1.06374	1.61737	−1.22811
124	0.52636	−0.51804	−1.00478	−0.61554
125	−0.63237	−0.45461	0.95248	0.33763
126	−0.66046	−0.0097	−0.33071	−0.72134
127	1.49548	2.40251	−0.15036	1.93301
129	−0.33541	−0.25214	−0.84392	0.80337
130	−0.93133	−0.63013	−0.20297	0.04377
131	−0.91628	−0.33208	0.35532	1.41089
132	0.95273	1.2581	−0.33872	0.0798
133	0.42804	0.73064	−0.97853	−0.74975
134	−1.70883	−0.18176	−0.94396	−0.05579

Table 5: Parameter definition.

Notation	Definition
T	Total travel time
$I_{\min}$	Minimum departure interval
t_{wf}	L1 and L2 passengers waiting time
q_1, q_2, q_3	Number of L1, L2, L3 passengers
f_{22}	L2 express train passenger ratio
C_{stop}	Parking surcharges
C_T	Passenger travel time cost
n_s	The number of stops of express train
H_f	Minimum departure interval between express train and following local train
A	Train capacity
$q_{\max}$	Maximum section flow
T_w	Total waiting time
t_{zhui}	Train tracking time
t_{ws}	L3 passengers waiting time
T_{in}	Total on-board time
f_{21}	L2 local train passenger ratio
c_s	Unit stopping cost
C_0	Unit time value
n_f	The number of stops of local train
H_s	Minimum departure interval between the local train and the following express train
h	Number of interchange stations
β	Load factor

Table 6: Notation for decision variables.

Notation	Definition
$a_s, s = 1, \ldots n$	1, if the train stops at station s; 0, otherwise
D_s	Pairs of local trains departing per unit time
D_f	Pairs of express trains departing per unit time

$$T_{in3} = \sum_{i=1}^{N-1} \sum_{j=2}^{N} q_{ij} t_{in}^{L3}. \quad (23)$$

Passengers' total on-board time

$$T_{in} = T_{in1} + T_{in2} + T_{in3}. \quad (24)$$

3.3.4. Objective 2. Minimum train operating costs

The cost changes brought about by this model only involve the additional cost of stops, so the second objective is expressed as follows:

$$\min C_{stop} = c_s \left(D_f n_s + D_s n_f \right). \quad (25)$$

The driving condition constraints are as follows:

$$a_1 = a_n = 1, \sum_{s=1}^{n} a_s < n, \tag{26}$$

$$D_f > 1, D_s > 1.1 \leq \frac{D_s}{D_f} \leq 2, \tag{27}$$

$$D_f + D_S = n, \tag{28}$$

$$D_{\min} \leq D_f + D_s \leq \frac{T}{I_{\min}}, \tag{29}$$

$$\frac{q_{\max}}{A \cdot \beta_{\max}} \leq D_f + D_s \leq \frac{q_{\max}}{A \cdot \beta_{\min}}, \tag{30}$$

$$2 + h < n_f < n - 2, \tag{31}$$

$$t_{\text{zhui}} \leq H_f \leq \frac{T}{2D_f} - t_{\text{zhui}}, \tag{32}$$

$$t_{\text{zhui}} \leq H_s \leq \frac{T}{2D_s} - t_{\text{zhui}}, \tag{33}$$

$$H_s D_s + H_f D_f \leq T. \tag{34}$$

Formula (26) represents the constraint on the stopping sequence of express trains; it stipulates that express trains must stop at the start and end stations and across one stop at the whole line at least. Formulas (27), (28), and (29) represent the constraint on the range of departure frequency; the departure frequency of the express train is greater than that of the local train. Formula (30) is the constraint of the section load rate; formula (31) is the constraint of the number of stops of the express train, and formulas (32), (33), and (34) are the constraints of the departure interval.

3.4. Model Solving. The multiple objectives of the optimization model are contradictory. Generally, one goal enhancement is accompanied by the performance degradation of another one, so no solution can make each goal reach the optimal solution, but a set of multiple optimal solutions, that is, the Pareto optimal solution set. In essence, the traditional solving methods convert the multiple objective functions by linear weighting into a single objective problem. The defect of this method is that the dimensions between the optimization objectives are difficult to be unified, and the value of the weight of each objective function remains to be discussed. Therefore, from the perspective that there are multiple satisfactory solutions for the subobjective solution of the multiobjective optimization model, it solves the Pareto solution set of the stopping scheme by using the binary coded express stopping scheme and the nondominated sorting NAGA-II algorithm with the elite strategy and analyzes each solution from the perspective of the potential stopping scheme and benefit. Finally, the optimal solution of the train stopping scheme and running pairs is obtained. The algorithm steps are as follows:

Step 1: Initialize the population and set the evolutionary algebra $Gen = 1$.

Step 2: Determine whether the first generation of subpopulation is generated; if it is generated, the evolutionary algebra $Gen = 2$; otherwise, the initial population is generated by nondominated ordering and selection, crossover, and mutation to generate the first generation of subpopulation, and the evolutionary algebra $Gen = 2$.

Step 3: Combine the parent population with the child population to form a new population.

Step 4: Determine whether a new parent population has been generated; if not, calculate the objective function of the individuals in the new population, and perform operations such as fast nondominated sorting, calculation of crowding degree, and elite strategy to generate a new parent population. Otherwise, proceed to step 5.

Step 5: Select, cross, and mutate the parent population to generate the offspring population.

Step 6: Determine whether Gen is equal to the largest evolutionary algebra; if not, then evolutionary algebra $Gen = Gen + 1$ and return to step 3. Otherwise, the run ends.

4. Case Study

4.1. Background. The total length of Changchun Metro Line 1 is 18.1 km, with 15 stations numbered 121 to 135. Stations 121 and 128 are transfer stations; Station 121 can transfer to Line 8, Station 128 can transfer to Line 2, Station 132 can transfer to Line 3, and there are external transportation hubs nearby: Changchun Expressway passenger station. Station 125 can transfer to Line 3, and Station 124 can transfer to Line 4 (transfer outside the station); the two stations are very close to each other and are connected to a large external transportation hub: Changchun Station. The specific geographical location of the route is shown in Figure 2.

4.2. Analysis of Passenger Flow Characteristics

4.2.1. Time Distribution of Passenger Flow. The maximum section passenger flow distribution of Changchun Rail Transit Line 1 is shown in Figure 3.

As can be seen from Figure 3, the maximum section passenger flow of this line presents a bimodal distribution, showing an obvious morning and evening peak at 7:00–9:00 and 17:00–19:00.

4.2.2. Section Distribution of Passenger Flow. The section distribution of passenger flow for Changchun Rail Transit Line 1 in the morning peak is shown in Figure 4.

(1) Direction unbalance coefficient

It can be seen from Figure 4 that in the morning peak hours, the distribution of section passenger flow up and down is “convex font,” that is, the section passenger flow at two ends of the line is small, and

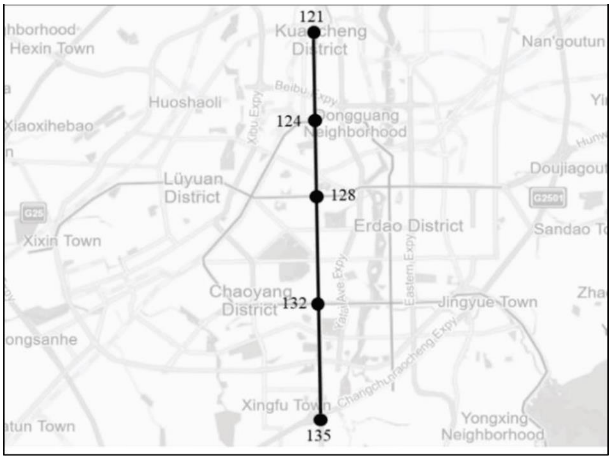

Figure 2: Diagram of line location information.

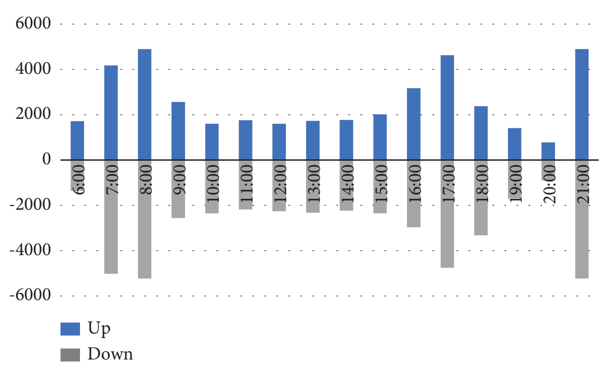

Figure 3: Maximum section passenger flow distribution.

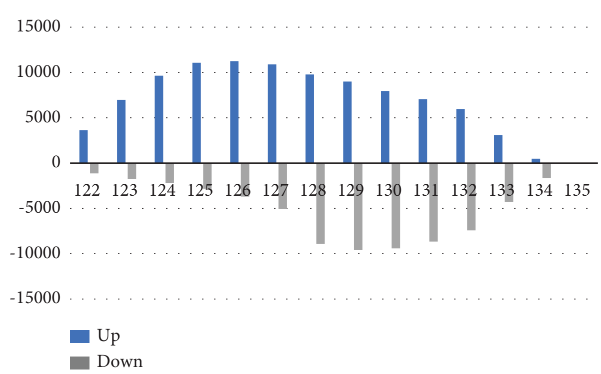

Figure 4: Cross-section passenger flow in the morning peak.

the section passenger flow in the middle is large. Overall, the passenger flow in the upward direction is slightly larger than that in the downward direction.

Passenger flow direction unbalance coefficient:

$$\lambda_1 = \frac{\max\left\{q_{\max}^{\text{up}}, q_{\max}^{\text{down}}\right\}}{\left(q_{\max}^{\text{up}} + q_{\max}^{\text{down}}\right)/2} = \frac{5627.5}{(5627.5 + 4709)/2} = 1.07. \tag{35}$$

The direction unbalance coefficient of this line is 1.07, and the closer it is to 1, the more balanced the distribution of upstream and downstream passenger flow, so the passenger flow of this line is more balanced in the direction.

(2) Section unbalance coefficient

The distribution of cross-section passenger flow on the station can be described by the unbalance coefficient of the passenger flow cross-section, as shown in Figure 5. The section unbalance coefficient of the passenger flow section exceeds 1.5 in both the upward and downward directions, and the passenger flow shows obvious fluctuation in space.

4.2.3. Distribution of Passenger Collector-Distributor Volume

(1) Passenger collector-distributor volume of each station:

The passenger flow of each station in the morning peak is shown in Figure 6.

As can be seen from Figure 6, at Stations 121, 122, 123, 124, and 134, the number of passengers getting on is much higher than that getting off, while at Stations 126, 127, and 129, the number of passengers getting off is much higher than that getting on, and the difference between the number of passengers getting on and off at Station 127 is particularly obvious. The number of passengers getting on and getting off at Station 128 is much higher than that at other stations. Combined with the analysis of the land use near the metro line, it can be preliminarily judged that Stations 127 and 128 are located in the city center, Station 127 is the destination of most commuter passengers, and Station 128 is also a transfer station, so there are both commuters and transfer passengers getting off at this station. Stations 121, 122, 123, and 134 are located in the suburbs, and the nearby areas are mostly residential land, so the number of passengers getting on is much higher than the number of passengers getting off.

(2) Passenger collector-distributor volume in different directions:

To further analyze the distribution of passengers getting on and getting off, the number of passengers in the upward and downward directions are counted separately, as shown in Figure 7.

As can be seen from Figure 7, in the upward direction, a large number of passengers get on the train at Stations 121, 122, 123, and 124, while few passengers get off. The situation is just the opposite at Stations 127, 132, and 133. Many passengers are getting on and off at Station 128, and there are almost no passengers getting on at stations south of Station 128. In the downward direction, a large number of passengers get on at Stations 132, 133, and 134, and a large number of passengers get off at Station 127. Similarly, a large

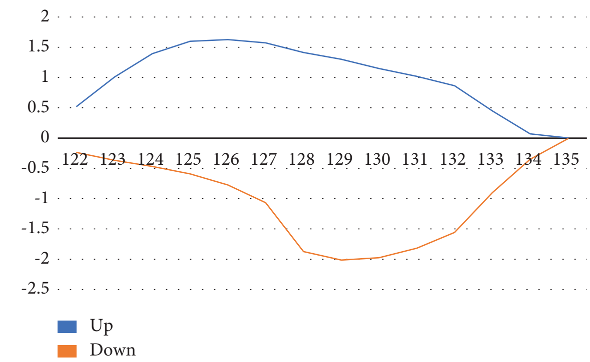

Figure 5: Passenger flow section unbalance coefficient.

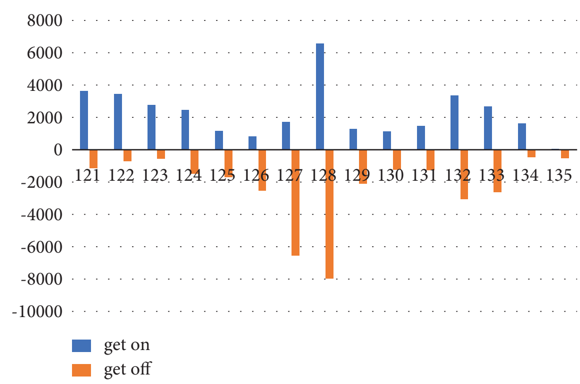

Figure 6: Morning peak passenger flow at each station.

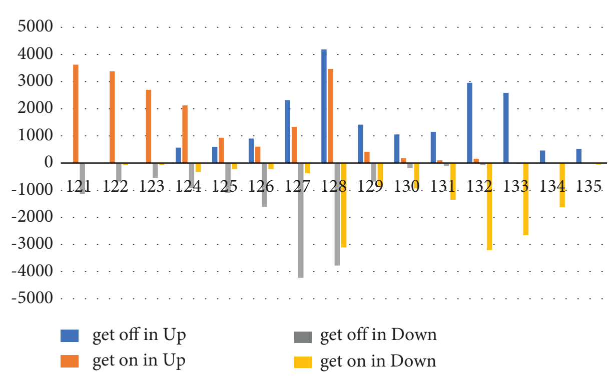

Figure 7: Morning peak passenger volume.

number of passengers get on and off at Station 128, and almost no passengers get off at stops south of Station 128. The most significant difference between the downward direction and the upward direction is that in the upward direction, there are still a large number of passengers getting off at stations south of Station 128, but in the downward direction, passengers are getting off at stations north of Station 128(except Station 127); there are very few passengers getting off at each station. This is because the station around the southern end of the metro line has both residential and office land, while the site around the northern end of the line is mostly residential land, and there are few office buildings.

4.2.4. Distance Distribution of Passenger Flow. The average station distance of Changchun Rail Transit Line 1 is about 1.2 km, the average passenger distance is 5.64 km, the average passenger distance of upward is 5.88 km, and the average passenger distance of downward is 5.56 km. The proportion of long-haul passengers is shown in Table 7.

All the above passenger flow indicators are given in Table 8.

Obviously, the passenger flow volume of each station of Changchun Rail Transit Line 1 is unbalanced, and there are more long-haul passengers; it is suitable for running express and local trains.

4.3. Optimization Results and Analysis. SPSS was used to perform cluster analysis on the standardized passenger flow data, residence index, and office index, and the cluster lineage diagram of the site was obtained, as shown in Figure 8.

According to the pedigree chart, the station can be divided into three categories and five subcategories, as given in Table 9.

The first type (Stations 122 and 123): these two stations have the highest residence index and the lowest office index among all the stations, indicating that the land around these sites is mostly residential land, and the passenger flow is relatively large. It can be inferred that the passenger flow of these sites during the morning peak period is mostly residents living around the site, so the stop is necessary at these two stations.

The second type: the second type of site has a low passenger flow, and the residence index and office index are in the middle level. Class 2.1 has Stations 124, 132, and 133, and the passenger flow of these stations is relatively large, and the office index is greater than the residence index, so there will be some passengers with the origin station (such as Stations 122 and 123) to the destination of Class 2.1 stations, and Stations 124 and 132 are transfer stations. So, this kind of station must be stopping stations. The common feature of Class 2.2 stations is that the passenger flow is small, so there is no need for the express train to stop. The residence index and office index around category stations (Stations 125 and 131) are high, but the passenger flow rate is small; the reason is that these areas are still in the development stage, so in the current train operation plan, such stations are designated as overtaking stations. The passenger flow of Class 2.2.2 is small, and the residence index and office index are not high, so the express train does not have to stop. It is worth mentioning that the distance between Stations 125 and 124 is only 500 m, and consecutive stops in short distance greatly affect the efficiency of subway operation, and Line 1 and Line 3 cannot realize intrastation transfer at Station 125, resulting in a small number of passengers at the station, so before realizing intrastation transfer between Line 1 and Line 3, it makes little sense for trains to stop at the station.

The third type (Station 127): Station 127 is divided into a separate category because the passenger flow in Station 127 is the largest; the office index is the highest among all stations, and the residence index is low, indicating that most of

Table 7: The proportion of long-haul passengers.

Item	Upward	Downward
Proportion of long-haul (>9 km) passengers (%)	37	14

Table 8: Passenger flow indicators.

Item	Upward in the morning peak	Downward in the morning peak
Direction unbalance coefficient	1.07	
Passenger flow section unbalance coefficient	1.62	2.01
Uneven coefficient of station distribution	3.04	3.4
Average passenger distance (km)	5.88	5.56

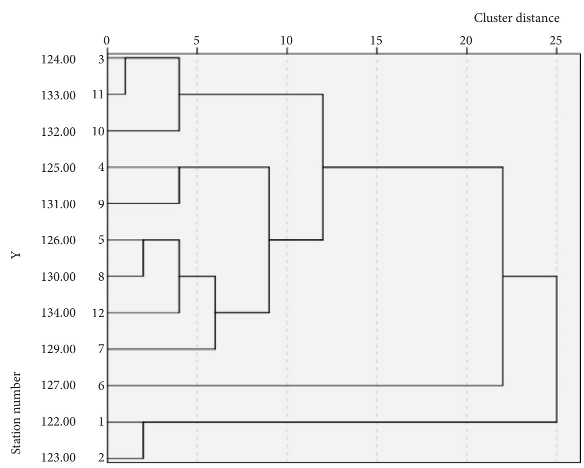

Figure 8: Station clustering pedigree diagram.

Table 9: Station classification table.

Category			Station number			
1	1.1		122	123		
2	2.1		124	132	133	
	2.2	2.2.1	125	131		
		2.2.2	126	130	134	129
3	3.1		127			

the land around such stations is office land. There is a large commuter flow, and there is a certain number of other types of passenger flow, so such stations must be stopping stations.

Based on the analysis of the passenger flow and clustering results, the potential stop stations and preliminary stopping plan are determined as follows: running express trains in the upward direction and running local trains in the downward direction. To sum up, the stopping scheme of fast and slow trains is determined as shown in Figure 9.

In the NSGA-II algorithm, the population size is 100, crossover probability = 0.8, mutation probability = 0.1, and the maximum number of iterations is 300. The model parameter values are given in Table 10. The schematic diagram of the Pareto solution set is shown in Figure 10. Four representative solutions selected from the Pareto solution set for comparison from the point of view of total travel time and operating costs, as shown in Table 11. 0 represents running slow train only, and the value in parentheses is the objective function value under the single data source: OD matrix. The timetable is given in Table 12, and the stopping scheme and the train diagram are shown in Figures 11 and 12.

The optimization results are analyzed as follows:

(1) Pareto solution 1. This scheme has the lowest number of stops, the lowest operating cost of the enterprise, and the highest total travel time of passengers, indicating that among the four schemes, this scheme has the lowest level of passenger service. If decision-makers give more importance to saving the operating cost of the enterprise, this scheme would be chosen.

(2) Pareto solution 2. In this scheme, the total travel time of passengers is the least, indicating that this scheme

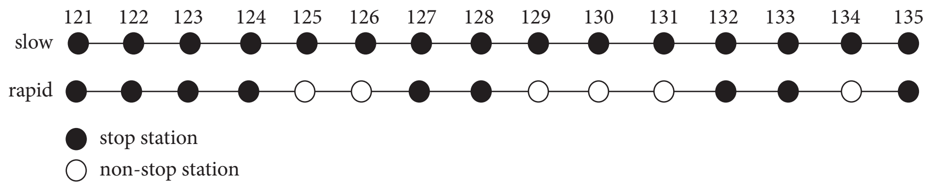

FIGURE 9: Schematic diagram of the potential train stopping scheme.

TABLE 10: Model parameter values.

Parameter	Values
The optimized period T	3600 s
Number of stations N	15
Unit stopping cost c_s	200 yuan/time
Stop time t_{stop}	35 s
Maximum running speed v_{max}	80 km/h
Start and stop additional time t_{qt}	25 s
Minimum train tracking interval t_{zhui}	90 s
Pairs of trains $D_f + D_s$	12 pairs
Unit time value C_0	0.013 yuan/s
Train capacity A	1888
Load factor β_{max}/β_{min}	0.8/0.2
Maximum section flow q_{max}	5627
Number of transfer stations h	4

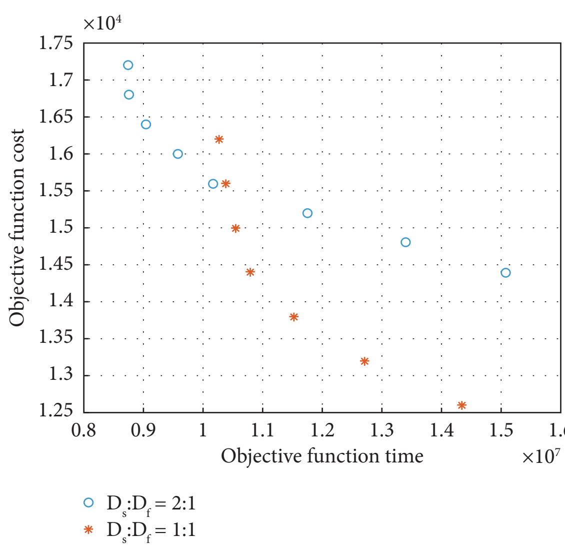

FIGURE 10: Convergence curve diagram.

TABLE 11: Pareto solution set comparison table.

Number	Pairs of express train (h)	Pairs of local train (h)	Total travel time (h)	Operating cost (yuan)	Save time (s)	Save cost (yuan)
0	—	12	3820.9	18000	—	—
1	4	8	2427.1 (2483.9)	17200 (17900)	1393.8	800
2	6	6	3782.7 (4409.3)	12600 (13800)	38.2	5400
3	4	8	2822.0 (3334.9)	15600 (16400)	998.9	2400
4	6	6	2996.5 (3289.7)	14400 (15700)	824.4	3600

Table 12: The train schedules.

Departure time	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135
1	7:33:00	7:34:54	7:36:53	7:38:46	7:40:09	7:41:58	7:43:48	7:45:46	7:47:58	7:50:01	7:51:55	7:53:35	7:55:43	7:58:22	8:00:11
2	7:36:00	7:37:54	7:39:52	7:41:46	7:43:09	7:44:58	7:46:48	7:48:46	7:50:58	7:53:01	7:54:55	7:56:35	7:58:43	8:01:22	8:03:11
3	7:45:00	7.46:54	7:48:53	7:50:46	7:51:09	7:52:58	7:53:48	7:55:46	7:56:58	7:58:01	7:58:55	8:00:35	8:02:43	8:04:22	8:06:11
4	7:48:00	7:49:54	7:51:52	7:53:46	7:55:09	7:56:58	7:58:48	8:00:46	8:02:58	8:05:01	8:06:55	8:07:35	8:10:43	8:13:22	8:15:11
5	7:51:00	7:52:54	7:54:52	7:56:46	7:58:09	7:59:58	8:01:48	8:03:46	8:05:58	8:07:01	8:09:55	8:11:35	8:13:43	8:16:22	8:17:11
6	8:00:00	8:01:54	8:03:52	8:05:46	8:06:09	8:06:58	8:08:48	8:10:46	8:11:58	8:13:01	8:13:55	8:15:35	8:17:43	8:19:22	8:21:11
7	8:03:00	8:04:54	8:06:52	8:08:46	8:10:09	8:11:58	8:13:48	8:15:46	8:17:58	8:20:01	8:21:55	8:23:35	8:25:43	8:28:22	8:30:11
8	8:06:00	8:07:54	8:09:53	8:11:46	8:13:09	8:14:58	8:16:48	8:18:46	8:20:58	8:23:01	8:24:55	8:26:35	8:28:43	8:31:22	8:33:11
9	8:15:00	8:16:54	8:18:52	8:20:46	8:21:09	8:21:58	8:23:48	8:25:46	8:26:58	8:28:01	8:28:55	8:30:35	8:32:43	8:34:22	8:36:11
10	8:18:00	8:19:54	8:21:53	8:23:46	8:25:09	8:26:58	8:28:4	8:30:46	8:32:58	8:35:01	8:36:55	8:38:35	8:40:43	8:43:22	8:45:11
11	8:21:00	8:22:54	8:24:52	8:26:46	8:28:09	8:29:58	8:31:48	8:33:46	8:35:58	8:38:01	8:39:55	8:41:35	8:43:43	8:46:22	8:48:11
12	8:30:00	8:31:54	8:33:53	8:35:46	8:36:09	8:36:58	8:38:48	8:40:46	8:41:58	8:43:01	8:43:55	8:45:35	8:47:43	8:49:22	8:51:11

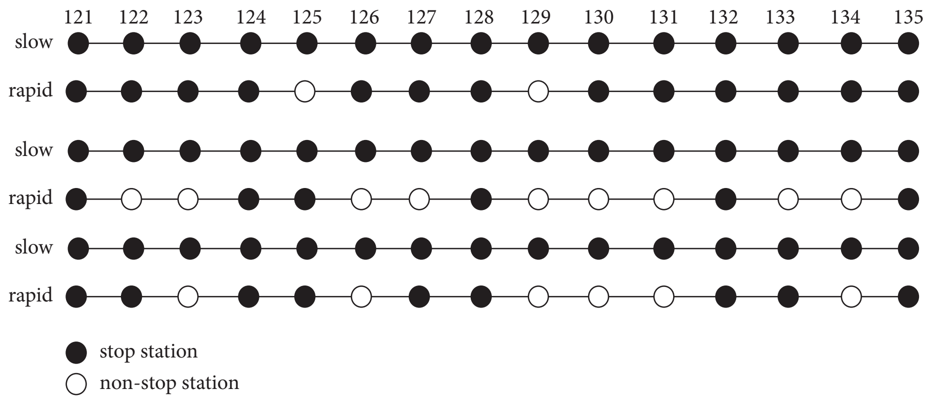

Figure 11: Schematic diagram of the stopping plan.

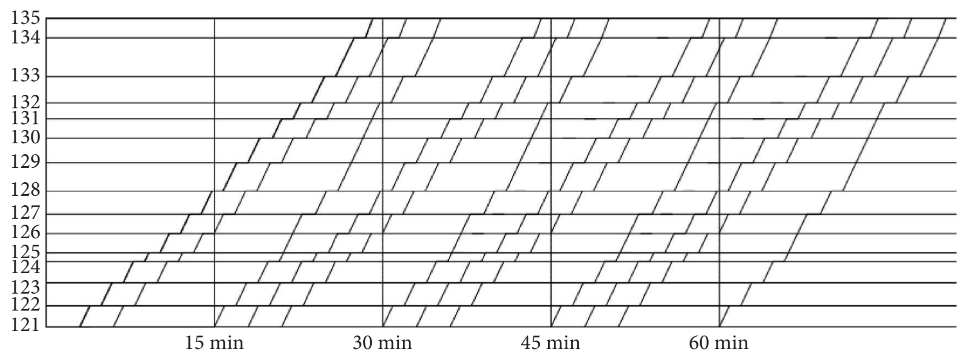

Figure 12: The train diagrams.

is more conducive to improving service quality of passengers. However, due to the large number of stops of express trains and the high operating cost of enterprises, this scheme would be chosen when decision-makers give more importance to the travel experience of passengers.

(3) Pareto solution 3. Compared with solutions 1 and 2, the scheme with the highest degree of overlap between the previous stopping sequence and the potential stopping scheme has a total passenger travel time greater than solution 2 and less than solution 1, and its operating cost is also between the two. Therefore, when decision-makers give more importance to the passenger service level, enterprise cost, and passenger travel experience, this scheme is more appropriate.

(4) Pareto solution 4. Similar to the solution 3 stopping scheme, this scheme is more appropriate when decision-makers value both passenger service level and enterprise cost, but value enterprise cost more.

5. Conclusion

In this paper, based on the multisource data of rail transit, we propose to optimize the design of the train stopping scheme and solve it by the NAGA-II algorithm, which improves the accuracy and real-time performance of the stopping scheme based on the existing research in this area. The model results are analyzed and compared from two aspects: passenger travel time and enterprise cost. A Pareto solution set is obtained for the best number of fast and slow trains and stopping schemes.

(1) Potential stop stations are determined based on land use nature, and the spatial weight function is introduced to obtain the land use index reflecting the characteristics of passenger flow. The stop plan could be adjusted according to land use planning and land development mode along the metro line in the future.

(2) The proposed model and algorithm can provide effective ideas for the preparation of rail transit train operation plans and guide passengers to travel reasonably; it has certain practical significance for the rail transit research.

This paper studies the parking scheme of fast and local trains in peak hours, assuming that the passenger flow obeys uniform distribution, and the peak coefficient of the passenger flow can be considered in further research. Besides, the integration of connected bus lines and passenger flow data as well as update frequency of data source can also be considered.

Data Availability

The statistic data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 71871103) and Scientific Research Project of Jilin Education Department in 2023 (JJKH20231189KJ).

References

[1] J. E. Cury, F. A. C. Gomide, and M. J. Mendes, "A methodology for generation of optimal schedules for an underground railway system," *IEEE Transactions on Automatic Control*, vol. 25, pp. 217–222, 1979.
[2] G. M. Smrek, "Urban transit: operations, planning and economics," *Operation*, vol. 45, no. 3, pp. 74-75, 2005.
[3] W. O. Assis and B. E. A. Milani, "Generation of optimal schedules for metro lines using model predictive control," *Automatica*, vol. 40, no. 8, pp. 1397–1404, 2004.
[4] Y. Wang, A. D'Ariano, J. Yin, L. Meng, T. Tang, and B. Ning, "Passenger demand-oriented train scheduling and rolling stock circulation planning for an urban rail transit line," *Transportation Research Part B: Methodological*, vol. 118, pp. 193–227, 2018.
[5] H. Niu, X. Zhou, and R. Gao, "Train scheduling for minimizing passenger waiting time with time-dependent demand and skip-stop patterns: nonlinear integer programming models with linear constraints," *Transportation Research Part B: Methodological*, vol. 76, pp. 117–135, 2015.
[6] P. Zhang, Z. Sun, and X. Liu, "Optimized skip-stop metro line operation using smart card data," *Journal of Advanced Transportation*, vol. 2017, pp. 1–17, 2017.
[7] Z. Cao, Z. Yuan, and D. Li, "Estimation method for a skip-stop operation strategy for urban rail transit in China," *Journal of Modern Transportation*, vol. 22, no. 3, pp. 174–182, 2014.
[8] Y. J. Lee, S. Shariat, and K. Choi, "Optimizing skip-stop rail transit stopping strategy using a Genetic Algorithm," *Journal of Public Transportation*, vol. 17, no. 2, pp. 135–164, 2014.
[9] R. Zhang, S. Yin, M. Ye, Z. Yang, and S. He, "A timetable optimization model for urban rail transit with express/local mode," *Journal of Advanced Transportation*, vol. 2021, pp. 1–15, 2021.
[10] L. Tang and X. Xu, "Optimization for operation scheme of express and local trains in suburban rail transit lines based on station classification and bi-level programming," *Journal of Rail Transport Planning & Management*, vol. 21, pp. 100283–100312, 2022.
[11] H. Niu, M. Chen, and M. Zhang, "Optimization theory and method of urban rail transit train running scheme," *China Railway Science*, vol. 32, no. 4, pp. 128–133, 2011.
[12] L. Deng, Q. Zeng, W. Gao, and W. Zhou, "Research on urban rail transit train operation scheme based on elastic demand," *Journal of the China Railway Society*, vol. 34, no. 12, pp. 16–25, 2012.
[13] Y. Zheng and W. Jin, "The elaboration of subway train operation scheme based on IC card data," *Journal of South China University of Technology*, vol. 43, no. 8, pp. 119–125, 2015.
[14] S. Zhao, J. Wu, Z. Li, and G. Meng, "Train operational plan optimization for urban rail transit lines considering circulation balance," *Sustainability*, vol. 14, no. 9, p. 5226, 2022.
[15] R. Yang, B. Han, Q. Zhang, Z. Han, and Y. Long, "Integrated optimization of train route plan and timetable with dynamic demand for the urban rail transit line," *Transport metrica B: Transport Dynamics*, vol. 11, no. 1, pp. 93–126, 2023.
[16] L. Liu, X. Ye, and B. Gu, "The revelation of the combined operation mode of Tokyo private railway fast and slow train to Shanghai metro line," *Urban Rail Transit Research*, vol. 11, pp. 38–41, 2006.
[17] Y. Sun, H. Shi, Y. Wang, and S. Chen, "Double-layer planning model of urban rail transit fast and slow vehicle operation scheme," *Journal of Transportation Systems Engineering and Information Technology*, vol. 18, no. 3, pp. 160–167, 2018.
[18] L. Zheng, R. Song, S. He, and H. Li, "Urban rail transit cross-station parking scheme optimization model and algorithm," *Railway Journal*, vol. 31, no. 6, pp. 1–8, 2009.